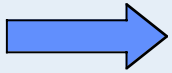
Welcome to HDF Workshop V

"Selected" ESDIS Status
What is HDF/HDF-EOS

Richard Ullman

ESDIS

Outline



- Welcome
- “Selected” ESDIS Status
- What is HDF/HDF-EOS

HDF & HDF-EOS Workshop 5

- **This is the fifth "annual" HDF workshop sponsored by ESDIS and NCSA.**
- **Combined this year with the Science Data Processing (SDP) workshop after our September meeting was canceled.**
- **We use this workshop to help guide our priorities for HDF and HDF-EOS development over the next year. Please make it a "workshop" by asking questions and expressing opinions.**
- **Consulting opportunities with NCSA and with ECS**
  - 2 lunches
  - After hours
  - Please sign up
- **Panel discussion tomorrow afternoon.**
- **Presentations from this workshop will be posted on the web on the HDF-EOS tools and information web site:**

http://hdfeos.gsfc.nasa.gov

Also on the Science Data Processing workshop page:

http://that.gsfc.nasa.gov/gss/workshop2002/workshop.html

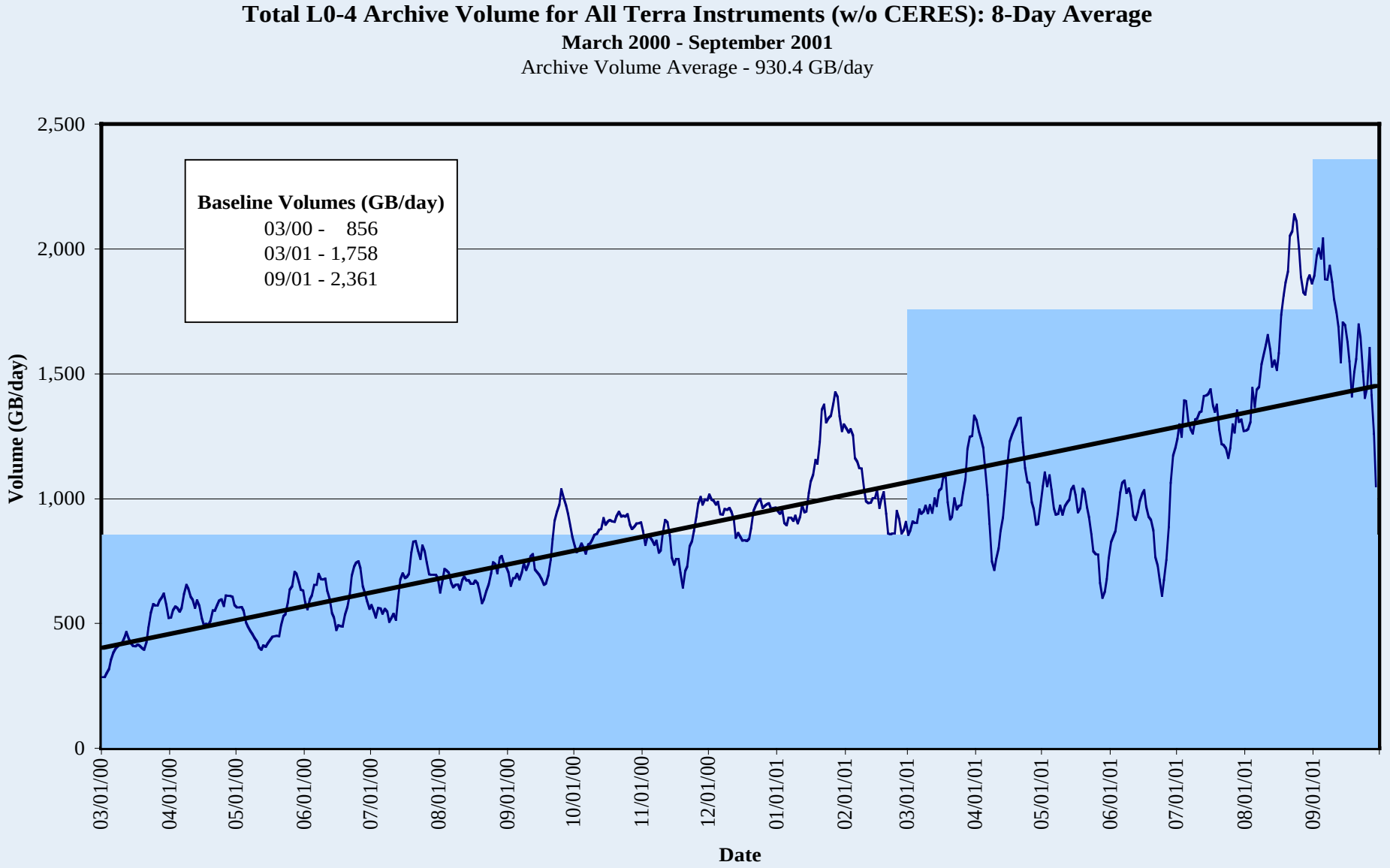
Outline

- Welcome
- "Selected" ESDIS Status
- What is HDF/HDF-EOS

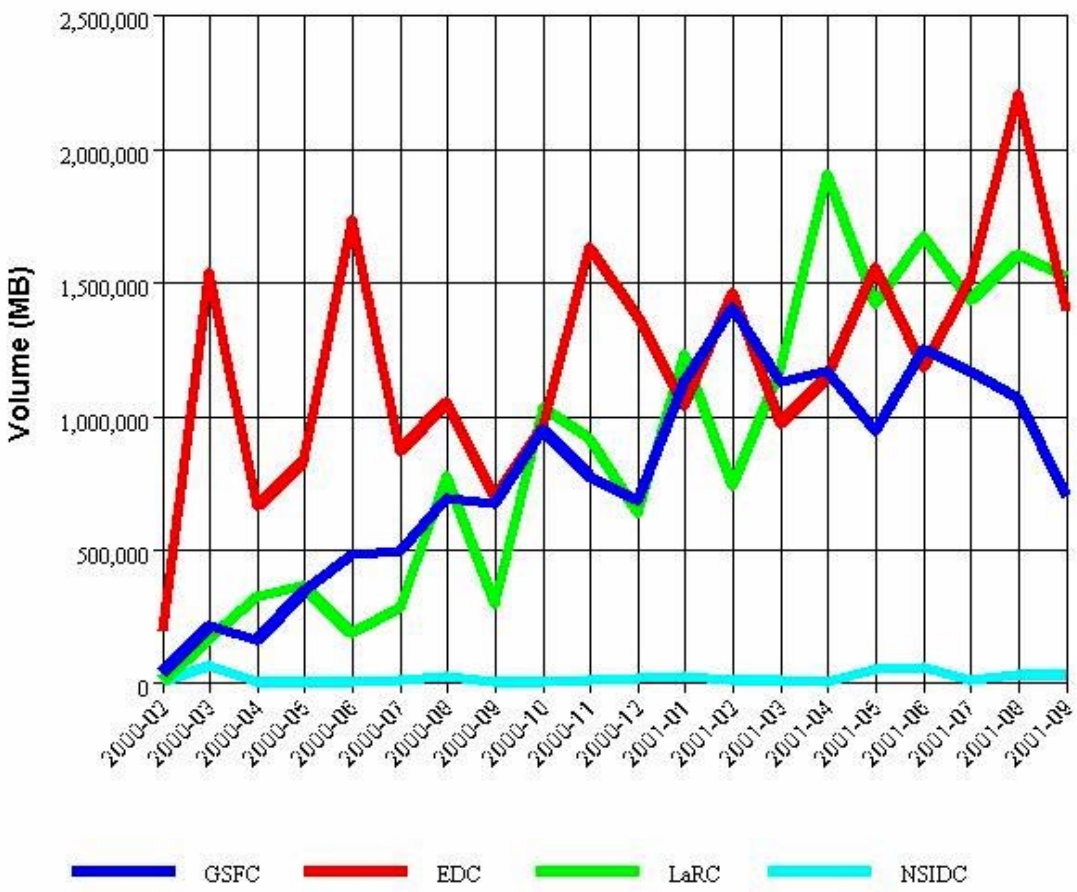
Data Availability

- **Since Terra launch, the EOSDIS science systems (ECS, DAACs, and SIPS) have processed and made available over 750 Terabytes Terra and Landsat 7 products.**
- **In many cases, validation of Terra products is continuing and care should be taken when using these data for science research.**
  - ASTER data, CERES data, and some MODIS products are at the "validated" state
- **ASTER**
  - Validated: Radiance at sensor, AST06 Decorrelation stretch, AST04 Brightness temperature
  - Provisional: AST09 Surface radiance-VNIR,SWIR, AST09T Surface radiance-TIR, AST07 Surface reflectance-VNIR,SWIR , AST08 Surface kinetic temperature, AST05 Surface emissivity
- **CERES "Edition 1" validated data released. "Edition 2" reprocessing expected to be completed June '02.**
- **MISR products upgraded from "beta" to "provisional":**
  - Geometric Parameters (from MISR PGE7)
  - L1B2 Terrain Radiance(a.k.a. GRP_TERRAIN_GM, MI1B2T) (from MISR PGE1)
  - L1B2 Ellipsoid Radiance
- **MODIS data reprocessing is in process, many datasets in "provisional status"**
- **MOPITT**
  - Limited amounts of Beta quality data are available from the LaRC DAAC.

Terra Archive Volume Against "Baseline Volumes"



EOS Data Distributed to End-Users (via EDG)



Future Planning - Data Pools

- **ECS DAACs are being augmented with a new delivery capability - Data Pools**
- **Data Pools are large, on-line, smart disk farms attached to ECS that allow users to quickly download key, popular data sets**
  - The Data Pool hardware and software will be installed at the DAACs in first quarter of FY02.
  - Data pools have metadata inventories to enable data location and filtering including personalized data view and to support DAAC controlled data residency.
  - Special services, including search and subsetting, will be available to users through the data pool interface (different from EDG).
    - User interface is via a "navigation" client based on GSFC DAAC-developed WHOM interface
  - Data pools are being procured using funds from the ECS "Synergy" earmark managed by NASA HQ/YO.
    - Operations concept for the data pools is being coordinated with YO (Gubbels) to ensure adequate support for YO applications requirements
    - DAAC Managers will work with their UWGs and YO to establish a data management plan for their Data Pool

Future Planning - EMD Contract

- **The ESDIS Project is preparing to procure an ESDIS Maintenance and Operations contractor**
- **ECS Maintenance and Development (EMD) Contract will be an Indefinite Delivery Indefinite Quantity (IDIQ) procurement**
  - Broad scope of potential work provides flexibility with minimum commitment
  - Specific work added by multiple Task Orders
  - 5 year overall period of performance; specific tasks define their own periods of performance (e.g., Task 1 will be 3 years)
- **Extended transition period between ECS and EMD designed to ensure transfer of knowledge of developed operational ECS to EMD winner.**
- **Target RFP release is Mid-March 2002.**

Community Advisory Process – DAAC UWG

- **The DAACs and the ESDIS Project continue to receive critical assessment and recommendations from the DAAC User Working Groups (UWGs)**
  - The UWGs are comprised of discipline and cross-discipline scientists and by applications researchers and data users.
  - The UWG Chairs have met through telecons to exchange ideas and lessons learned regarding the UWG process.
    - Initiated and chaired by the ESDIS Project Scientist
  - UWGs meet every 6-9 months and are always looking for new members to augment their panels.

Community Advisory Process – DAWG

- **The ESDIS Project has formed a special Data Access Working Group (DAWG)**
  - The DAWG is comprised of representatives from the DAACs, the Science Working Group on Data (SWGD), the instrument science teams, the SIPSs, and the general user community
  - The DAWG is addressing operations concepts and system inefficiencies that may be limiting access to EOS data at the DAACs
    - Key problem areas have been identified and prioritized for work off.
    - Where problems are already known (through existing NCRs) the priority for work off of those NCRs has been raised.
    - Other, new NCRs, have been written against the system to improve access throughput.
    - Identifying new capabilities that would support refined data access operations concepts
  - The DAWG process has been very effective, due in large part to the "targeted" focus of the working group
    - Need and effectiveness of focused "tiger teams" is a key lesson learned for NewDISS

Community Advisory Process – SWGD

- **The Science Working Group on Data (SWGD) is comprised of representatives from the EOS instrument teams (primarily Terra), the DAACs, and technical experts from the ESDIS project.**
- **The SWGD is an advocate for EOS science processing in an advisory role in helping ESDIS prioritize use of enhancement resources**
  - The SWGD has been looking at throughput requirements for the EOS science data processing and distribution based on measured performance of systems and algorithm software.
  - SWGD has made numerous recommendations to the Project for hardware enhancement to improve throughput in key, target areas.
  - The SWGD as begun looking at other data center issues including distribution and tool availability.

Outline

- Welcome
- “Selected” ESDIS Status
- What is HDF/HDF-EOS

What is HDF?

- **A container for storing scientific data**
  - A portable format and a portable library
- **Stores images, multidimensional arrays, tables, etc.**
- **Emphasis on storage and I/O efficiency**
- **Free and commercial software support**
- **Emphasis on standards**
- **Users from many engineering and scientific fields**

What is HDF5?

- **New format(s) and library (c.1998)**
  - Format is not compatible with HDF4 library
  - Primary Objects:
    - Groups
    - Datasets
  - Secondary Objects:
    - Datatypes
    - Dataspaces
  - Additional means to organize data
    - Attributes
    - Sharable objects
    - Storage and access properties

- **New Demands**
  - Bigger, faster machines and storage systems
    - massive parallelism, teraflop speeds
    - parallel file systems, terabyte storage
  - Greater complexity
    - complex data structures
    - complex subsetting
  - Emphasis on remote & distributed access

- **New HDF5 Features**
  - More scalable
    - Larger arrays and files
    - More objects
  - Improved data model
    - New datatypes
    - Single comprehensive dataset object
  - Improved software
    - More flexible, robust library
    - More flexible API
    - More I/O options

What is HDF-EOS?

- **To share files, users must organize them similarly.**
- **HDF user groups create standard profiles**
  - Ways to organize data in HDF files.
  - Metadata
  - API and library
- **HDF-EOS is the standard profile of HDF for EOS standard products**
  - Swath
  - Grid
  - Point
  - Profile (Atmospheric profile)

HDF-EOS lessons learned

- **Ten years since first "trade study" for Earth Systems Science data format standard.**
  - "Baseline standard" decision September 1993(nine years)
  - First library was released June 1996 (six years)
  - HDF-EOS 5 released spring 2001 (one year)
- **The concept of a "profile" for interchange of files among members of a producer/user community is reasonable.**
- **It has been difficult for HDF-EOS development to quickly respond to community feedback.**
- **Standard still allows a great deal of flexibility.**
  - This flexibility permits mission or instrument science teams to create products with the subtlety that they require.
  - Flexibility also means that general purpose (cross-product) end-user tools are difficult to support.
- **"Standards are difficult to negotiate but essential for success."**

  - Dolly Perkins (SDP Workshop 2/27/02)